

# Analysis of morphology-based features for classification of crop and weeds in precision agriculture

Petra Bosilj, Tom Duckett, and Grzegorz Cielniak  
{pbosilj,tduckett,gcielniak}@lincoln.ac.uk

**Abstract**—Determining the types of vegetation present in an image is a core step in many precision agriculture tasks. In this paper, we focus on pixel-based approaches for classification of crops versus weeds, especially for complex cases involving overlapping plants and partial occlusion. We examine the benefits of multi-scale and content-driven morphology-based descriptors called Attribute Profiles. These are compared to state-of-the-art keypoint descriptors with a fixed neighbourhood previously used in precision agriculture, namely Histograms of Oriented Gradients and Local Binary Patterns. The proposed classification technique is especially advantageous when coupled with morphology-based segmentation on a max-tree structure, as the same representation can be re-used for feature extraction. The robustness of the approach is demonstrated by an experimental evaluation on two datasets with different crop types, while being able to provide descriptors at a higher resolution. The proposed approach compared favourably to state-of-the-art approaches without an increase in computational complexity, while being able to provide descriptors at a higher resolution.

## I. INTRODUCTION

Precision agriculture concerns the use of monitoring and intervention techniques to improve efficiency, especially to reduce or avoid the use of pesticides and herbicides, and boost crop yield. This can be achieved by mounting vision systems on robotic systems such as agricultural implements, ground robots or drones to monitor the crop and recommend treatments where needed. In order to perform tasks such as selective spraying or mechanical weeding, a number of approaches to discriminate value crop from weeds have been developed [1], [2], [3], [4].

Mathematical morphology [5] offers a versatile framework to perform multi-scale spatial analysis of image content in various domains. Attribute morphology, relying on hierarchical image representations [6], offers ways to capture different features such as texture, size and shape across multiple scales at a pixel [7], region [8], [9] and image [10] level.

The contributions of this paper are as follows:

- A novel pixel-based crop versus weed discrimination approach based on attribute profiles [7], [11].
- Improvement over state-of-the-art methods based on HOG [12], [13], [14] and LBP [15], [16], [4] descriptors on a standard public dataset [17] and our own challenging data set.
- Adapting the automatic parameter selection for attribute profiles [11] for application to sequences of images.

## II. RELATED WORK

An important goal of image processing in precision agriculture is to determine the position and type of vegetation. This information enables selective per-plant operations, such as weeding, spraying or harvesting, on a large scale.

The first step is usually soil removal [18], segmenting out the background information and identifying the positions of all vegetation in the image. The input are colour images represented in different colour spaces, or index-images calculated from various spectra [19], [20], [18]. Applicable approaches, including global and adaptive thresholding [18], local region-based approaches [9] and machine learning techniques [2], vary in their complexity and precision. After background removal, the remaining (vegetation) pixels are then presented to a classifier to determine the plant type, or more generally distinguish between value crops and weeds. The classifier needs to robustly handle varying lighting and weather conditions, plant overlap and image artefacts.

The two main approaches to plant classification are region-based [4], [9] and pixel-wise [3], [16], which can be combined to obtain the benefits of both [4]. In region-based approaches, each connected region from the segmentation output is represented by a set of descriptors. These work well when not too many of the plants overlap (e.g. at early stages of growth), but cannot cope with mixed vegetation regions. To deal with overlap and partial occlusion, a pixel-based classifier reaches a decision for each pixel based on its local neighbourhood. The classification can be based directly on the features calculated on the image [16], [9], or further statistical features extracted from feature images [3], [4]. A recent approach applying deep learning methods to augmented input data also reported good results [21].

Pixel-wise approaches come at the cost of reduced speed as they have to classify thousands of individual pixels rather than dozens of regions. The number of pixels can be reduced by combining the two approaches and removing the regions classified with high confidence. Keypoint descriptors are calculated based on a local neighbourhood determined by the image resolution, distance of the camera to the crop, and expected size and type of plants. Further computational savings can be achieved by classifying on a sparse sampling grid based on the minimum expected plant size [3], [4].

Mathematical morphology [5] traditionally comprises tools which enable the scale-analysis of images. The rigid structuring element is replaced by a content-driven neighbourhood defined by nested regions in multi-scale hierarchi-

<sup>1</sup>Lincoln Centre for Autonomous Systems, School of Computer Science, University of Lincoln, LN6 7TS Lincoln, UK

cal image representations [22], [6]. As decisions are reached based on different attributes calculated for the regions of the hierarchy, such tools are considered as a part of *attribute morphology* [23]. Attribute Profiles [7] are multi-scale spatial-spectral pixel descriptors from the framework of attribute morphology, applied mostly in classification of satellite data [24]. Following recent advances removing the need for manual parameter tuning [11], we explore the performance of this descriptor for plant type classification in precision agriculture, and compare with state-of-the-art descriptors based on Histograms of Oriented Gradients [12] and Local Binary Patterns [15].

### III. METHODOLOGY

While multi-class classification between weed types is possible, here we designed a binary classifier to distinguish between value crops and weeds. Individual points are classified using a descriptor vector representing their local neighbourhood, which should have high discriminative power while remaining invariant under image transformations (e.g. intensity scaling, rescaling or rotation). We use the Random Forest (RF) classifier [25], an ensemble approach relying on the principle that a group of “weak learners” (classifiers capturing only some aspect of the data) can together form a “strong learner”, which better approximates the desired decision function. The weak learners used in RF are decision trees [26], each trained on a subset of the training data and splitting the dataset based on a random subset of the prediction variables. The RF output is the most frequent output of its decision trees. Hereafter, we explain the descriptors used for plant type classification.

#### A. Attribute Profiles

Attribute Profiles (AP) [7] are multi-scale pixel description tools, constructed by successively filtering the image with increasingly coarser filters. They rely on Attribute Filters (AF), connected morphological filters dealing directly with connected components (CC) of the image instead of pixels. AP based on area can be explained in a similar way to granulometries [5]: in each step, all image components below a certain size are removed from the image with the size being increased for every step, which can be compared to sieving the image through a stack of sieves with increasing mesh size. All the resulting images are stacked, and a single descriptor is formed by concatenating the different pixel intensities through the stack. The process is similar for any attribute, removing components below a certain attribute value in each step.

More formally, given a monochannel image  $f : E \rightarrow \mathbb{Z}, E \subseteq \mathbb{Z}^2$ , its upper-level set at level  $t$  is defined as  $\mathcal{L}^t = \{f \geq t\}$  with  $t \in \mathbb{Z}$  (resp. lower-level set  $\mathcal{L}_t = \{f \leq t\}$ ). A collection of all upper-level sets corresponds to a set of images obtained when thresholding at all possible intensity values. The image is typically evaluated with 4-connectivity, enabling the distinction between different CC within a level set, called peak components. AF are realized by evaluating a logical predicate on each peak component and preserving the

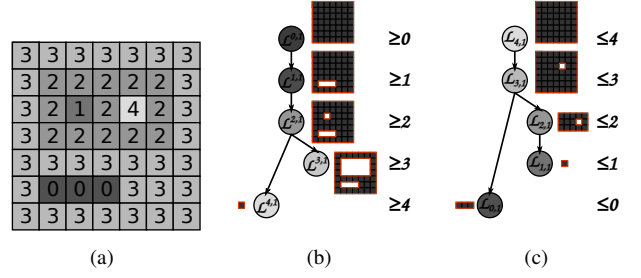


Fig. 1. The max-tree for the shape in (a) is displayed in (b), and the min-tree in (c). The regions corresponding to the nodes of the trees are displayed besides them, with the level sets indicated inside the nodes.

selected components. To construct a sequence of increasingly coarser filters, the logical predicate  $T_\kappa^\alpha$  is typically defined through an attribute  $\alpha$  and a threshold value  $\kappa$ , and evaluated by comparing the value of  $\alpha$  for a selected CC to  $\kappa$  (e.g.  $T_{500}^{\text{area}}$  accepts CC with area  $> 500$  pixels).

A sequence of ordered logical predicates is typically used to define the AF as attribute thinnings and thickenings [23]. For  $T^\alpha$  and  $L$  thresholds  $\{\kappa_i\}$  with  $1 \leq i \leq L$  with corresponding attribute thinnings  $\gamma^{\kappa_i}$  and thickenings  $\phi^{\kappa_i}$ , the AP of an image  $f$  is defined as:

$$AP(f) = \{\phi^{\kappa_L}(f), \phi^{\kappa_{L-1}}(f), \dots, \phi^{\kappa_1}(f), f, \gamma^{\kappa_1}(f), \dots, \gamma^{\kappa_{L-1}}(f), \gamma^{\kappa_L}(f)\}. \quad (1)$$

The descriptor for a single pixel consists of the values it obtains after each AF. The descriptor length is  $1 + L$  for one-sided AP as used in this paper (only thinning or thickening; corresponding to the top or bottom half of Eq. (1) and  $1 + 2L$  in case of two-sided AP (full Eq. (1)).

The recent popularity of AP is due to their efficient implementation using a max-tree and min-tree [22]. These dual representations structure the inclusion relations between peak components of the lower (resp. upper) level sets  $\mathcal{L}_t$ , which are nested for increasing (resp. decreasing) values of  $t$ . The monochannel image is fully represented, with bright structures modelled by the min-tree, and dark by the max-tree (see examples in Fig. 1). Multiple AF are calculated from the same hierarchy as node or branch removal operations.

The attribute  $\alpha$  and thresholds  $\{\kappa_i\}$  are needed to fully define an AP. We use three different attributes, measuring different characteristics in the image:

- **Area (A)** is the size of the region in pixels, producing AP sensitive to objects at multiple scales.
- **Standard deviation (S)** of the pixel intensity values, providing a measure of the texture of the region.
- **Moment of inertia (I)** of the region is equivalent to the first moment invariant of Hu [27] calculated as  $I = \eta_{2,0} + \eta_{0,2}$ , where  $\eta_{0,2}$  and  $\eta_{2,0}$  are central normalized image moments, representing the shape of the region.

The AP thresholds depend on image scale and content, and can be determined automatically from the granulometric curve (GC) of the image [11]. For each threshold  $\kappa$ , the GC measures the amount of image content (in terms of

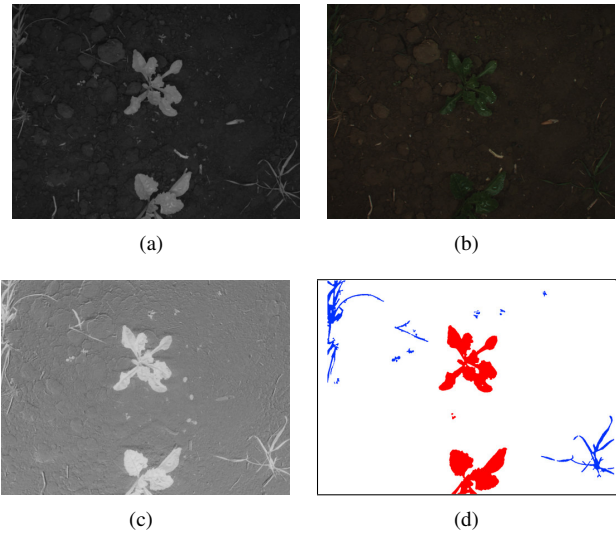


Fig. 2. Examples of NIR and RGB images from the Sugar Beets 2016 dataset are shown in (a) and (b) respectively. The resulting NDVI image is shown in (c), while the ground truth is displayed in (d). The sugar beets are shown in ‘red’, with the weeds in ‘blue’.

Dataset	#img	image size	#samples /img (avg)	class distrib.	crop rows	plants /img (avg)
<i>Sugar Beets 2016</i>	280	1296 × 966	75k (6% of image)	C - 68% W - 32%	1	C - 2.25 W - 8
<i>Carrots 2017</i>	20	2428 × 1985	1015k (21% of image)	C - 35% W - 65%	4 × 2	C - 88 W - 85.6

TABLE I

DETAILS OF THE DATASETS USED FOR EVALUATING THE PROPOSED APPROACH. IN THE CLASS DISTRIBUTION COLUMN ‘C’ STANDS FOR ‘CROP’ AND ‘W’ FOR ‘WEEDS’.

number of changed pixels, sum of their intensities, or number of changed CC) removed by the corresponding AF. It is constructed directly from the hierarchy without performing AF, and the set of optimal AP thresholds is obtained as breakpoints between segments when approximating the GC with piecewise linear regression [28].

As one of the contributions of this paper, we propose to approximate a composite GC obtained by summing  $N > 1$  selected GC according to the chosen measure. This way, the optimal threshold can be determined for a multi-image dataset typical in precision agriculture, in contrast to remote sensing where both the training and testing samples are drawn from the same image (e.g. [7], [24], [29]). The number of thresholds is also determined automatically, and we obtained  $L < 10$  in all of our experiments.

### B. Local Binary Patterns

Local Binary Patterns (LBP) are texture descriptors calculated based on the local neighbourhood of each pixel, by comparing the pixel intensity to its neighbours [15]. As in the original paper, we use the 8-neighbourhood, and encode the sign of the comparisons at each pixel position as an 8-bit binary value. The descriptor for the whole image is then

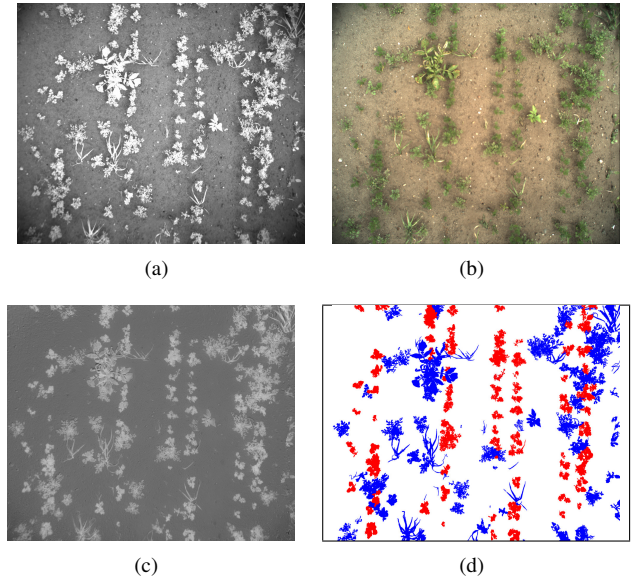


Fig. 3. Examples of NIR and RGB images from the Carrots 2017 dataset are shown in (a) and (b) respectively. The resulting NDVI image is shown in (c), while the ground truth is displayed in (d). The carrots are shown in ‘red’, with the weeds in ‘blue’.

typically produced as a histogram of all the LBP values. A further improvement can be introduced by distinguishing so-called ‘uniform’ from ‘non-uniform’ patterns. A binary pattern is considered uniform, if, when considered as circular it contains at most two bitwise transitions  $0 \rightarrow 1$  or  $1 \rightarrow 0$  (e.g. patterns 00111000 and 11100000 are both uniform with 2 transitions, while the pattern 00100110 is non-uniform with 4 transitions). The histogram will contain upwards of 70% uniform patterns, so separately binning only the uniform values without distinguishing between the non-uniform ones improves the rotation invariance [30]. As there are 18 distinct uniform patterns, the length of the descriptor is 19.

Since we are interested in classifying pixels or grid-points, the LBP for the whole image are calculated only once and the descriptor histograms are accumulated across local image patches of size  $M \times M$ .

### C. Histograms of Oriented Gradients

Histograms of Oriented Gradients (HOG) descriptors are based on the distribution of intensity gradients or edge orientations. They have been successfully used to encode shape in plant leaf classification. The image is divided into cells of size  $c \times c$  and histograms of gradient directions with  $d$  bins are calculated for each cell. To achieve a better invariance to illumination changes, the cells are further grouped into overlapping blocks containing  $b \times b$  cells (choosing an odd  $b$  ensures that they are centred in the defined cells), and the histograms are contrast-normalized across each block.

Typically, both in its original application to human detection [12] and in plant leaf classification [13], the HOG descriptors are used to describe whole objects, and template matching with a sliding window can be employed for object detection. This is done by concatenating all the histograms of all the defined blocks, where each object-image or window

is resized to ensure the descriptor is calculated on the same number of cells and blocks. Since we are interested in classifying pixels and their local neighbourhood we calculate one histogram per block, obtained by concatenating and contrast-normalizing only the histograms of the cells within that block. The length of each cell histogram is  $d$ , while the whole block descriptor has a length of  $d \times b^2$ .

#### IV. DATA AND EXPERIMENTAL SETUP

##### A. Image acquisition

We apply our methods to multi-spectral vegetation data combining RGB colour information with a near infra-red (NIR) channel. A combination of red (R) and NIR information, called Normalized Difference Vegetation Index (NDVI), is a good indication of photosynthetic plant activity [19]. The NDVI image can be calculated as:

$$NDVI = \frac{NIR - R}{NIR + R}. \quad (2)$$

To validate our approach, we evaluate the proposed technique on two sets of NDVI data, summarized in Table I.

The *Sugar Beets 2016* dataset contains RGBN images of crop rows sown with a broad-leaved crop, collected under controlled lighting using the BoniRob agricultural robot [17]. The JAI 130-GE camera used to collect this dataset uses a splitting prism resulting in perfectly aligned RGBN images and consequently no errors in NDVI calculation. It was mounted inside an opaque shroud under the robot chassis, providing independence from natural light. The resulting images have a size of  $1296 \times 966$  pixels with spatial resolution of 3 px/mm corresponding to a field patch with approximate dimensions  $24 \times 31$  cm. The dataset provides the per-pixel labels for 280 images used in the experiments, distinguishing between sugar beet crop and many types of weeds. For our experiments, the ground truth was changed so that all the weed vegetation belongs to a single class. Example images from this dataset are shown in Fig. 2.

The *Carrots 2017* dataset was collected by the authors in fields in Lincolnshire, UK in June 2017 using a two-camera setup with RGB and NIR cameras (Teledyne DALSA Genie Nano) mounted 5 cm apart. The images were obtained using a manually pulled cart over a carrot field, and no lighting control mechanisms. The crop shown is under substantial weed pressure and has similar appearance to weed. Since the optical centres of the two cameras are not aligned, obtaining the NDVI images requires an additional registration step which can result in parallax errors, especially on tall plants. The produced images have a high resolution of  $2464 \times 2056$  pixels (dictated by the camera parameters), corresponding to a spatial resolution of 2.5 px/mm for a  $100 \times 85$  cm patch of ground (a few pixels are lost later in the registration step). In comparison to the *Sugar Beets 2016* dataset, the images from *Carrots 2017* correspond to a 10 times larger surface area and contain 35 times as many crop plants and 10 times as many weed plants on average. To generate the ground truth, the images were manually annotated into weed and crop structures (with examples shown in Fig. 3).

Dataset	A	I	S
SB 2016	141, 464, 2767	0.271, 0.436, 0.817	6, 12
CA 2017	221, 2764, 15007	0.249, 0.341, 0.546	3

TABLE II

ATTRIBUTE PROFILE THRESHOLDS SELECTED BY AUTOMATIC THRESHOLD SELECTION. ‘SB 2016’ STANDS FOR ‘SUGAR BEETS 2016’ DATASET, AND ‘CA 2017’ FOR ‘CARROTS 2017’ DATASET. ‘A’ STANDS FOR ‘AREA’ ATTRIBUTE, ‘I’ FOR ‘MOMENTS OF INERTIA’ AND ‘S’ FOR ‘STANDARD DEVIATION’.

##### B. Experimental setup

The RF used for plant type classification used 10 decision trees. In order to test the performance of the classifier on both datasets, we perform  $k$ -fold cross validation (with  $k = 10$ ) by dividing the dataset images into  $k$  equally sized sets.  $k - 1$  sets of images are used to train the classifier, which are further divided internally into a training and validation set, and the performance is evaluated on the remaining set.

The automatic threshold selection needed for AP calculation was done on a random subset containing 5% of the larger *Sugar Beets 2016* dataset, and on half of the images from the smaller *Carrots 2017* dataset. Only the max-tree was used and one-sided AP calculated, as the NDVI images ensure vegetation will always be bright on a darker background. The measure used for the granulometric curve was the number of CC modified by each AF threshold. The results of automatic threshold selection for AP are summarized in Table II, with selected AP components for each attribute shown in Fig. 4. As the AP calculation is done by consecutively filtering the image max-tree representation, it is simultaneously calculated at all pixel positions.

The sampling grid size chosen is  $8 \times 8$  pixels. This determines the cell size for HOG descriptors as  $c = 8$ . For reasons of comparison with [4] who also evaluate on the *Sugar Beets 2016* dataset using the local neighbourhood of  $80 \times 80$  pixels, we chose HOG block size  $b = 5$  resulting in a similar local neighbourhood of  $72 \times 72$ . We chose the same neighbourhood size  $M = 72$  for the LBP descriptors. As the AP descriptors already include neighbourhood information at various scales at each pixel, no aggregation across any further local neighbourhood is done and only one AP corresponding to each grid centre is used for classification.

##### C. Performance evaluation

To ensure a fair comparison between classification methods, several performance measures were used. The accuracy (Acc) is simply the proportion of samples (pixels) that were correctly classified. However, in the case of unbalanced distributions, the accuracy itself does not give sufficient information about the performance of a classifier without knowing the expected probability of a chance agreement between the classifier and the ground truth  $\text{Acc}_{\text{chance}}$ . For this reason, we use Cohen’s Kappa coefficient [31] to measure the agreement between the classifier and ground truth:

$$\kappa = \frac{\text{Acc} - \text{Acc}_{\text{chance}}}{1 - \text{Acc}_{\text{chance}}}. \quad (3)$$

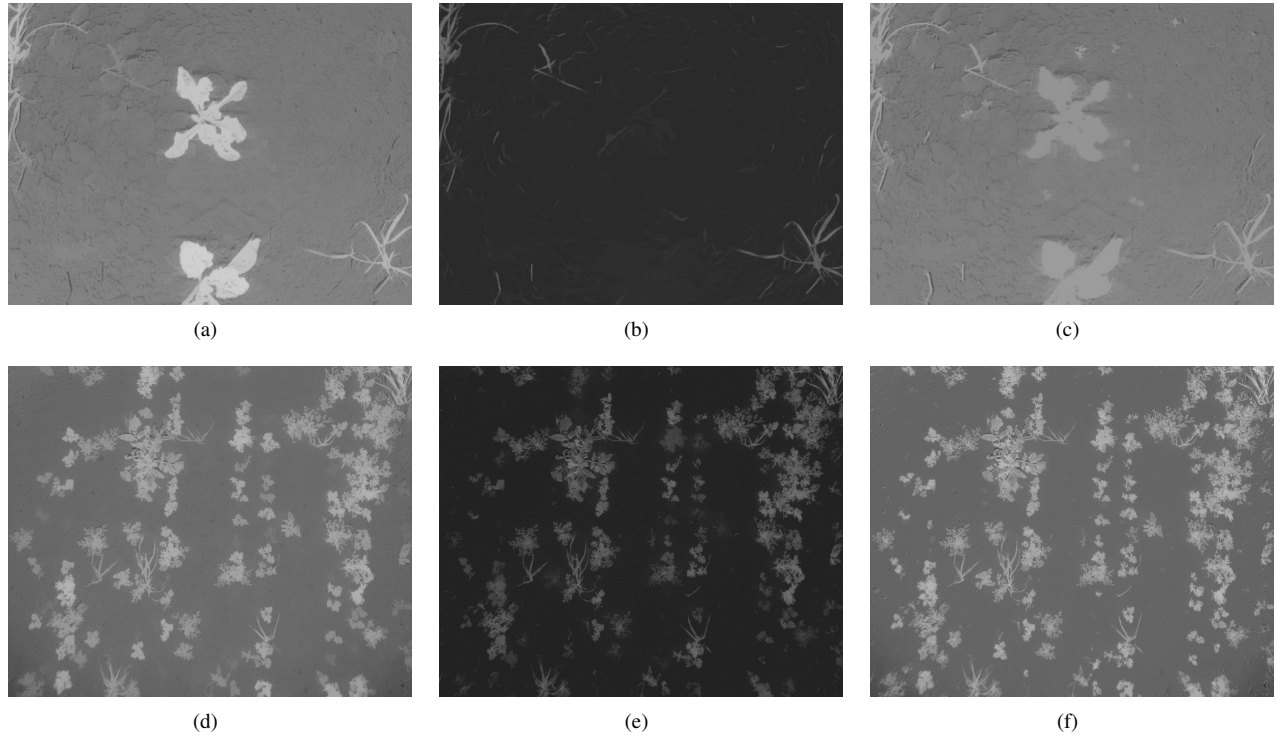


Fig. 4. Example of AP for different attributes (contrast of the components may change when filtering with shape or texture attributes). For the *Sugar Beets 2016* dataset: ‘area’ ( $\kappa = 2767$ ) (a), ‘moment of inertia’ ( $\kappa = 0.436$ ) (b) and ‘standard deviation’ ( $\kappa = 12$ ) (c). For the *Carrots 2017* dataset: ‘area’ ( $\kappa = 2764$ ) (d), ‘moment of inertia’ ( $\kappa = 0.249$ ) (e) and ‘standard deviation’ ( $\kappa = 3$ ) (f).

Descriptor (len)	no positional information						positional information					
	Crop		Weed		$\kappa$	Acc[%]	Crop		Weed		$\kappa$	Acc[%]
	p[%]	r[%]	p[%]	r[%]			p[%]	r[%]	p[%]	r[%]		
<i>Sugar Beets 2016</i>												
position (1)							85.79	94.14	83.92	66.23	0.64	85.32
HOG (200)	79.40	93.53	76.33	46.23	0.45	78.82	85.02	94.91	84.81	62.97	0.62	84.98
LBP (18)	84.28	91.55	76.86	62.18	0.57	82.41	89.56	94.58	86.30	75.58	0.73	88.67
HOG+LBP (218)	86.64	<b>94.98</b>	85.91	67.65	0.66	86.46	88.20	<b>95.24</b>	<b>87.23</b>	71.85	0.70	87.95
AP:A (4)	84.22	90.27	75.05	63.37	0.56	81.77	90.16	92.89	83.52	78.05	0.72	88.20
AP:S (3)	89.37	85.03	70.66	78.11	0.61	82.84	91.27	90.01	79.00	81.36	0.71	87.28
AP:I (4)	84.70	83.75	65.64	67.23	0.51	78.53	90.09	91.55	81.03	78.19	0.70	87.33
AP:I+S (6)	<b>90.19</b>	87.13	74.03	79.48	0.65	84.71	92.14	91.60	82.04	83.08	0.74	88.91
AP:A+I (7)	86.57	91.68	79.34	69.20	0.63	84.58	90.30	94.21	86.17	78.08	0.74	89.12
AP:A+S (6)	89.80	90.90	79.75	77.63	0.69	86.71	<b>92.14</b>	93.67	85.77	<b>82.68</b>	0.77	90.20
AP:A+I+S (9)	90.17	92.46	<b>82.71</b>	<b>78.17</b>	<b>0.72</b>	<b>87.94</b>	91.93	94.30	86.92	82.08	<b>0.78</b>	<b>90.44</b>
<i>Carrots 2017</i>												
position (1)							47.90	21.47	67.23	<b>87.33</b>	0.10	64.18
HOG (200)	44.10	38.64	68.17	72.85	0.12	60.65	45.28	40.75	68.88	72.70	0.14	61.31
LBP (18)	50.78	49.71	72.45	73.30	0.23	64.88	53.51	52.97	74.08	74.49	0.28	66.82
HOG+LBP (218)	50.65	47.54	71.88	74.32	0.22	64.77	51.75	48.77	72.48	74.80	0.24	65.51
AP:A (4)	<b>59.62</b>	47.64	74.40	82.51	0.32	70.25	57.04	56.67	76.59	76.85	0.34	69.76
AP:S (2)	56.93	32.40	70.29	<b>86.71</b>	0.21	67.62	48.00	44.25	71.00	74.01	0.19	63.55
AP:I (4)	47.19	39.64	69.89	75.95	0.16	63.18	50.37	43.73	71.53	76.64	0.21	65.07
AP:I+S (5)	48.88	39.06	70.20	77.85	0.18	64.21	50.97	45.75	72.13	76.14	0.22	65.46
AP:A+I (7)	57.09	<b>51.72</b>	75.09	78.92	0.31	69.36	<b>57.84</b>	54.13	75.97	78.61	<b>0.33</b>	<b>70.00</b>
AP:A+S (5)	57.25	51.03	74.93	79.34	0.31	69.39	56.54	<b>57.65</b>	<b>76.79</b>	75.97	<b>0.33</b>	69.53
AP:A+I+S (8)	58.57	51.70	<b>75.38</b>	80.17	<b>0.33</b>	<b>70.16</b>	57.70	54.48	76.04	78.35	<b>0.33</b>	69.96

TABLE III

RESULTS OF CLASSIFICATION ON A  $8 \times 8$  SAMPLING GRID. ‘P’ STANDS FOR PRECISION, ‘R’ FOR RECALL, ‘ACC’ SIGNIFIES ACCURACY. ‘HOG’ STANDS FOR HISTOGRAMS OF ORIENTED GRADIENTS, ‘LBP’ FOR LOCAL BINARY PATTERNS AND ‘AP’ FOR ATTRIBUTE PROFILES (‘A’ FOR ‘AREA’ ATTRIBUTE, ‘I’ FOR ‘MOMENT OF INERTIA’ AND ‘S’ FOR ‘STANDARD DEVIATION’). THE DESCRIPTOR LENGTH (‘LEN’) IS INDICATED IN BRACKETS.



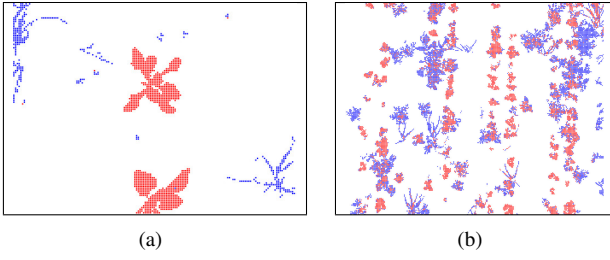


Fig. 5. Examples of classification output. The classification result for the *Sugar Beets 2016* image shown in Fig. 2 is shown in (a). The result for the *Carrots 2017* image shown in Fig. 3 is displayed in (b).

To further examine the performance of the classifier we also calculate precision and recall measures for each class. Precision indicates the chance that a sample classified into a class is correctly classified, and is calculated as:

$$p = \frac{TP}{TP + FP}, \quad (4)$$

while recall, or specificity, refers to the percentage of relevant samples of a class assigned to that class, expressed as:

$$r = \frac{TP}{TP + FN}. \quad (5)$$

where TP, FP and FN are the number of true positive, false positive and false negative samples, respectively.

## V. RESULTS

The comparison of classification methods for both datasets is shown in Table III, and for two example images in Fig. 5. We show the performance for all the examined descriptors, as well as the combinations of the classical descriptors and AP calculated for different attributes. Additionally, we show the performance when the horizontal image coordinate of the sample in the image was included in the descriptor. This corresponds to positional information related to crop row position, and shows an improvement in the classification results (an even better improvement could be achieved by estimating the crop row position [32] and including the distance of the sample to the crop row as a feature). The Bayesian confidence intervals, computed for all considered measures as the standard deviation of the Beta distribution posteriors over the belief in the accuracy assuming uniform priors [33], confirm the significance of all the results as displayed up to two decimal places.

The best classification rates on both datasets, both in terms of accuracy and the  $\kappa$  measure were achieved by attribute profile descriptors. More specifically, the combinations of ‘area’ and ‘standard deviation’ attributes and the combination of all three attributes showed the best and close second-best performance on both datasets. We show the feature importance calculated as Mean Decrease in Impurity [25] for the combination of all three attributes (before adding positional information) on both datasets in Fig. 6.

On the *Sugar Beets 2016* dataset we achieve excellent agreement with the ground truth according to the  $\kappa$  coefficient. As the crops are planted in a single row, the provided  $x$ -coordinate is more easily exploitable. The large

differences in scene complexity and content between datasets (cf. Table I) contribute to the lower performance of all the descriptors on the *Carrots 2017* dataset. Additionally, due to the skew of multiple crop rows present, the improvement achieved by including positional information is smaller. We have noticed that the difficult cases correspond to large weed clusters (typically grass), with some misclassifications along the edges of the crop class, especially in patches of mixed vegetation. Further improvement to the results could be achieved with spatial smoothing and interpolation as well as a combination with a region based approach [3], [4].

We also compare our results to the recent deep learning approach for plant classification [21] evaluated on *Sugar Beets 2016* dataset. We report an increase in precision for both ‘crop’ and ‘weed’ classes (of +6% and +20% respectively) with a minimal decrease in recall (−3% for both classes), suggesting a similar performance on the ‘crop’ class and an improvement on the ‘weed’ class for our pipeline. We note that in [21], the soil was treated as a separate ‘ground’ class, negatively impacting measurements for the two smaller classes. A fair comparison between the two approaches would require a standardized experimental setup.

We measured the execution times of descriptor calculation on the sugar beets images on an Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz. For the purpose of comparison, the descriptors were calculated for the whole image instead of just the vegetation patches. The LBP and HOG descriptors were calculated on a sampling grid, while the AP descriptors were calculated on whole images. We obtained similar execution times of 1.9s per image for HOG descriptors, 1.8s for LBP descriptors and 1.7 – 2.0s for AP descriptors depending on the attribute (of which 1.3s for hierarchy calculation and the rest for AF operations on the hierarchy). The computational complexity of AP calculation is linear in the number of image pixels once the thresholds are determined, comprising two linear operations of building the max-tree and performing AF. Calculating the thresholds is also linear in the total number of pixels used (i.e. image size  $\times$  number of images), consisting of building the max-tree and GC for each of the images, merging the GC and performing time series segmentation [28].

The advantage of AP descriptors relies on the local neighbourhood being inferred from the image instead of being fixed, which additionally enables to obtain them at any resolution with no additional cost. They allow classification based on multiple characteristics (descriptors based on size, shape, texture and any other attributes can be obtained simultaneously) with a minimal increase in computational complexity, while most classical descriptors focus on encoding only one trait (e.g. texture for LBP descriptors). This is additionally achieved with significantly shorter descriptors than for the competing methods examined. Finally, the difference in importance of the ‘standard deviation’ feature on *Sugar Beets 2016* (cf. Fig. 6) as well as the performance of the LBP descriptors suggests that texture is much more important in this dataset than for *Carrots 2017*.

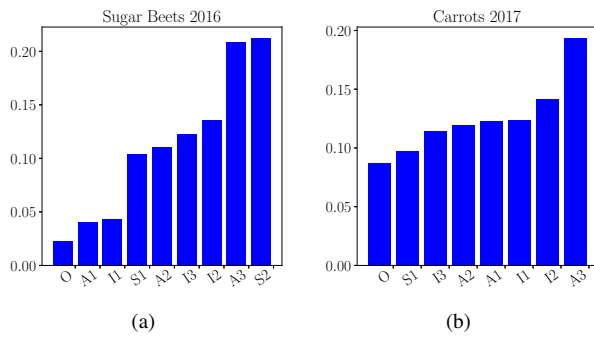


Fig. 6. Feature importance based on the coefficients of the RF classifier for the *Sugar Beets 2016* dataset is shown in (a) and for the *Carrots 2017* dataset in (b), as relative participation of each feature in the final decision, such that their sum is 1. The label 'O' stands for the intensity value of the original image, included in the descriptors, while the rest are the components of 'area' (A), 'moments of inertia' (I) and 'standard deviation' (S) profiles.

## VI. CONCLUSIONS AND FUTURE WORK

This paper presented a pixel-wise plant-type classification system based on attribute profiles, providing multi-scale and multi-attribute morphological pixel descriptors. We show a significant improvement over state-of-the-art descriptors as well as excellent classification rates on two datasets for different crops. Additionally, we adapted the automatic parameter tuning method for this descriptor [11] to make it suitable for multi-image datasets. We plan to further investigate attribute morphology as a versatile and multiscale framework for precision agriculture. More complex variants of AP descriptors [29] based on different hierarchies [6] could improve pixel-based classification. Further improvements could be achieved by combination with region-based morphological segmentation and classification [9], with an additional benefit of reusing the hierarchical image representation, the most computationally expensive step of both approaches, for segmentation as well as feature extraction.

## REFERENCES

- [1] J. Hemming and T. Rath, "Computer-vision-based weed identification under field conditions using controlled lighting," *J. Agricultural Engineering Research*, vol. 78, no. 3, pp. 233–243, 2001.
- [2] J. M. Guerrero, G. Pajares, M. Montalvo, J. Romeo, and M. Guijarro, "Support vector machines for crop/weeds identification in maize fields," *Expert Systems with Applications*, vol. 39, no. 12, pp. 11 149–11 155, 2012.
- [3] S. Haug, A. Michaels, P. Biber, and J. Ostermann, "Plant classification system for crop/weed discrimination without segmentation," in *Proc. WACV*, 2014, pp. 1142–1149.
- [4] P. Lottes, M. Hörferlin, S. Sander, and C. Stachniss, "Effective vision-based classification for separating sugar beets and weeds for precision farming," *J. Field Robotics*, vol. 34, no. 6, pp. 1160–1178, 2017.
- [5] P. Soille, *Morphological image analysis: principles and applications*. Springer Science and Business Media, 2013.
- [6] P. Bosilj, E. Kijak, and S. Lefèvre, "Partition and inclusion hierarchies of images: A comprehensive survey," *J. Imaging*, vol. 4, no. 2, p. 33, 2018.
- [7] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, 2010.
- [8] P. Bosilj, E. Aptoula, S. Lefèvre, and E. Kijak, "Retrieval of remote sensing images with pattern spectra descriptors," *ISPRS Int. J. Geo-Information*, vol. 5, no. 12, p. 228, 2016.
- [9] P. Bosilj, T. Duckett, and C. G., "Connected attribute morphology for unified vegetation segmentation and classification in precision agriculture," *Computers in Industry, Special Issue on Machine Vision for Outdoor Environments*, vol. 98, pp. 226–240, 2018.
- [10] E. Aptoula, "Remote sensing image retrieval with global morphological texture descriptors," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 3023–3034, 2014.
- [11] G. Cavallaro, N. Falco, M. Dalla Mura, and J. A. Benediktsson, "Automatic attribute profiles," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1859–1872, 2017.
- [12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. CVPR*, vol. 1, 2005, pp. 886–893.
- [13] X.-Y. Xiao, R. Hu, S.-W. Zhang, and X.-F. Wang, "HOG-based approach for leaf classification," in *Proc. ICIC*, 2010, pp. 149–155.
- [14] D. G. Tsolakidis, D. I. Kosmopoulos, and G. Papadourakis, "Plant leaf recognition using zernike moments and histogram of oriented gradients," in *Hellenic Conf. Artificial Intell.*, 2014, pp. 406–417.
- [15] D.-C. He and L. Wang, "Texture unit, texture spectrum, and texture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 4, pp. 509–512, 1990.
- [16] A. Binch and C. Fox, "Controlled comparison of machine vision algorithms for rumex and urtica detection in grassland," *Computers and Electronics in Agriculture*, vol. 140, pp. 123–138, 2017.
- [17] N. Chebrolu, P. Lottes, A. Schaefer, W. Winterhalter, W. Burgard, and C. Stachniss, "Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields," *Int. J. Robotics Research*, 2017.
- [18] E. Hamuda, M. Glavin, and E. Jones, "A survey of image processing techniques for plant extraction and segmentation in the field," *Computers and Electronics in Agriculture*, vol. 125, pp. 184–199, 2016.
- [19] J. Rouse Jr, R. Haas, J. Schell, and D. Deering, "Monitoring vegetation systems in the Great Plains with ERTS," *NASA Special Publication-351*, 1974, presented at Third ERTS Symposium.
- [20] D. Woebbecke, G. Meyer, K. Von Bargen, and D. Mortensen, "Color indices for weed identification under various soil, residue, and lighting conditions," *Trans. American Soc. Agricultural and Biological Engineers*, vol. 38, no. 1, pp. 259–269, 1995.
- [21] A. Milioto, P. Lottes, and C. Stachniss, "Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs," in *Proc. ICRA*, 2018.
- [22] P. Salembier, A. Oliveras, and L. Garrido, "Antiextensive connected operators for image and sequence processing," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 555–570, 1998.
- [23] E. J. Breen and R. Jones, "Attribute openings, thinnings, and granulometries," *Computer Vision and Image Understanding*, vol. 64, no. 3, pp. 377–389, 1996.
- [24] B. Song, J. Li, M. Dalla Mura, P. Li, A. Plaza, J. M. Bioucas-Dias, J. A. Benediktsson, and J. Chanussot, "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, 2014.
- [25] L. Breiman, J. Friedman, C. Stone, and R. Olshen, *Classification and Regression Trees*. Taylor and Francis, 1984.
- [26] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 3, pp. 660–674, 1991.
- [27] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.
- [28] D. Lemire, "A better alternative to piecewise linear time series segmentation," in *Proc. SIAM Int. Conf. Data Mining*, 2007, pp. 545–550.
- [29] G. Cavallaro, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Extended self-dual attribute profiles for the classification of hyperspectral images," *IEEE Geosci. Remote Sens. Letters*, vol. 12, no. 8, pp. 1690–1694, 2015.
- [30] T. Ojala, M. Pietikainen, and T. Maenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. PAMI*, vol. 24, no. 7, pp. 971–987, 2002.
- [31] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.
- [32] X. P. Burgos-Artizzu, A. Ribeiro, M. Guijarro, and G. Pajares, "Real-time image processing for crop/weed discrimination in maize fields," *Computers and Electronics in Agriculture*, vol. 75, no. 2, pp. 337–346, 2011.
- [33] J. M. Bernardo and A. F. M. Smith, "Bayesian theory," 2001.